

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number  
**WO 01/91110 A1**

(51) International Patent Classification<sup>7</sup>: **G10L 15/26**

(21) International Application Number: **PCT/EP01/05945**

(22) International Filing Date: 23 May 2001 (23.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
00401434.6 23 May 2000 (23.05.2000) EP

(71) Applicant (or all designated States except US): **THOMSON LICENSING S.A.** [FR/FR]; 46, quai Alphonse Le Gallo, F-92100 Boulogne Billancourt (FR).

(72) Inventors: and

(75) Inventors/Applicants (for US only): **TAZINE, Nour-Edine** [FR/FR]; 22, rue des Clos Blancs, F-35530 Noyal

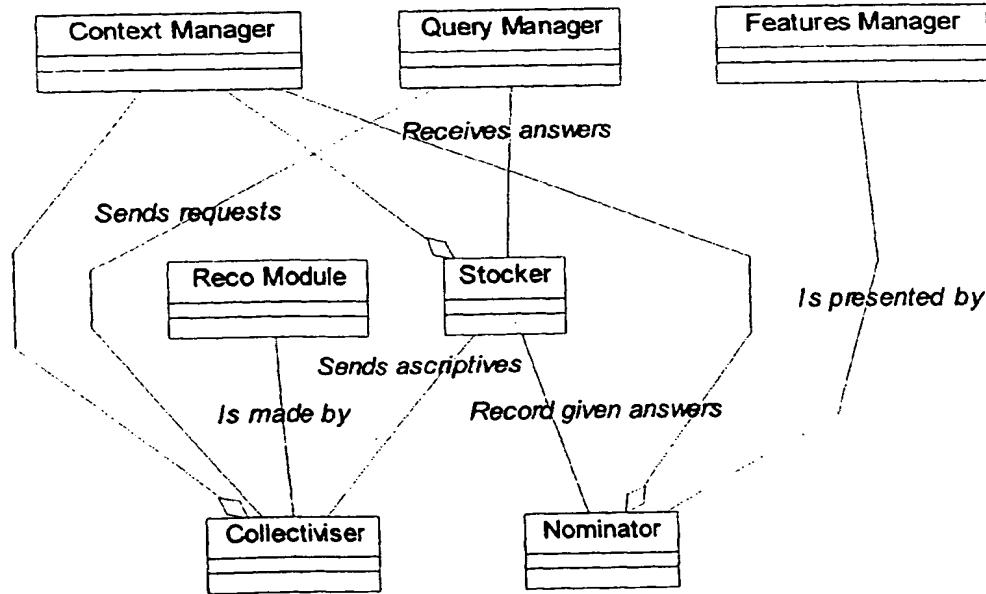
sur Vilaine (FR). **SOUFFLET, Frédéric** [FR/FR]; 18, rue de Normandie, F-35410 Chateaugiron (FR). **QUERRE, Gouven** [FR/FR]; 13, rue de la Crois des Forges, F-35530 Noyal sur Vilaine (FR). **LE HUITOUZE, Serge** [FR/FR]; 24, rue Bertrand Robidou, F-35000 Rennes (FR). **DE-LAUNAY, Christophe** [FR/FR]; 11, rue Pierre Varin de la Brumelière, F-35700 Rennes (FR). **JOUET, Pierrick** [FR/FR]; 27, boulevard de Lattre de Tassigny, F-35000 Rennes (FR). **ORLAC, Izabela** [FR/FR]; Le Champ Michel, F-35530 Noyal sur Vilaine (FR).

(74) Agent: **KERBER, Thierry**; Thomson Multimedia, 46, quai Alphonse le Gallo, F-92648 Boulogne Billancourt Cedex (FR).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) Title: STAND-ALONE DEVICE COMPRISING A VOICE RECOGNITION SYSTEM



WO 01/91110 A1

(57) Abstract: The invention concerns a stand-alone device comprising a voice recognition system and a mass storage device. The device comprises moreover a natural language processing, at least one semantic network which defines a domain, a database containing information with attributes. The natural language processing receives recognized command of a user and sends a request to the semantic network. The semantic network is constituted as a graph defining the elements of the domain and the links between these elements. The semantic network searches the answer to the request in the domain.



(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— with international search report

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

### Stand-alone device comprising a voice recognition system

The present invention concerns a voice-controlled device for the  
5 home, comprising a flexible voice-controlled user interface.

The object of the invention is a stand-alone device containing a voice recognition system, a mass storage device, characterized in that it comprises moreover a natural language processing, at least one semantic  
10 network which defines a domain, a database containing information with attributes, the natural language processing receiving recognized command of an user and sending a request to the semantic network, the semantic network being constituted as a graph defining the elements of the domain and the links between these elements, the semantic network searching the  
15 answer to the request in the domain.

Other characteristics and advantages of the invention will appear through the description of a preferred embodiment of the invention. This embodiment will be described in relation with the drawings among which:

20 Figure 1 is an illustration representing two possible appearances of the device.

Figure 2 is an illustration detailing the external features of a device of figure 1.

Figure 3 is a UML diagram of a context manager interface.

25 Figure 4 is a diagram of a collectiviser structure of the context manager of figure 3.

Figure 5 is a diagram of a stocker structure of the context manager of figure 3.

30 Figure 6 is a diagram of a nominator structure of the context manager of figure 3.

CONFIRMATION COPY

Figure 7 is a block diagram of a semantic network representing the knowledge in one particular domain, in this case for an Electronic Program Guide (EPG) application.

5 Figure 8 is a block diagram of a semantic network for a device command and control application.

Figure 9 is a diagram of a first basic relation in a semantic network.

Figure 10 is a diagram of a second basic relation in a semantic network.

10 Figure 11 is a diagram of a third basic relation in a semantic network.

Figure 12 is a diagram of a variant of relation in a semantic network.

Figure 13 is a diagram of a 'role' relation.

15 Figure 14 is a diagram of the iterative steps required to create a new context according to the embodiment.

Figure 15 is a flowchart of a tool used to create a new context.

Figure 16 is diagram of different steps used to generate language models for different languages.

20 Figure 17 is a diagram of a tool software architecture.

### 1.1. Introduction and link between Home Assistant and the voice-based EPG

In our former project "voice-based EPG", we have demonstrated, 25 by the building of a prototype, that a voice-operated application for a complex albeit restricted domain is feasible.

However, we also have noticed that the development of such an application from scratch is quite costly.

Building on top of this experiment, we propose to define and build 30 a framework to rationalize and ease the development of voice-controlled applications.

Central to this framework is the architectural split between generic components implementing the common tasks of a voice-based application (voice signal processing, speech recognition, text to speech, speaker authentication, ...) and pluggable modules for tasks specific data.

5 This will lead to a much shorter development time for a particular voice-operated application, because specific modules will be more high-level and generic components will be reused. This separation will also open the way to multiple applications being simultaneously active.

The Home Assistant (HA) is an implementation of this framework.

10 The HA will be a physical device able to run many voice-operated applications and to dynamically load/unload them from distant servers.

## 1.2. Introduction

Home assistant is a stand-alone device you can talk to spontaneously, almost as you would do with a human being. It responds in the same way, through a text to speech module, or it may display information on a screen. Home assistant can be moveable, it can work alone. It contains communication means to communicate with a network and download information.

20 With such a simple description, it can be thought of as a robot, and to avoid confusions, it is important to nail down the main differences between Home Assistant and apparently similar devices from other companies.

## 25 1.3. Features of Home Assistant

Let's start by describing the physical embodiment of HA. Figure 1 shows a possible embodiment of a home assistant. Figure 2 shows some of its externally visible features.

HA is a voice-operated device, basically consisting of a display, a  
30 microphone and loudspeakers.

Weighing a few kilograms, it can be easily moved by hand from place to place, but is not designed to be portable. Thus, it will presumably be somewhere in the living-room.

Two-way vocal interaction with the HA will be possible by 5 speaking right in front of it at a distance of 1 or 2 meters, and also everywhere in the home, through small, remote-control-like, devices with microphones and loudspeakers that will be lying on docking stations placed in most rooms.

The more salient features of HA will be:

- 10 • a display to present results to the user or animate HA's face;
- Text-To-Speech technology to allow distant interactions through wireless devices;
- high-quality audio system, to allow for audio CD and MP3 playing;
- 15 • speaker recognition, for greater customization of the interaction;
- memory of past interaction, for a more natural dialog;
- connection to various digital links, which allows:
- easy download of new applications for the HA via dedicated servers,
- 20 • very natural interaction to access data on the internet,
- control of every digital device present at home (IEEE 1394 link),
- distant interaction with HA through classical telephony.

25

### 1.3.1. The notion of knowledge modules

The HA's knowledge is organized into independent modules. Many modules may coexist simultaneously in the HA and new ones may be downloaded at any moment, increasing its "intelligence" (i.e. its ability to 30 understand discourse domains) accordingly.

There are technical and economic reasons for that:

- to keep individual module to a manageable size, thus allowing parallel development by different teams;
- to allow early availability of the product, through careful choice of target application domains;

5           • to shorten time to market for new, trendy, application domains;

- to allow new features to be integrated at any time, thus avoiding the premature obsolescence of the product;
- to allow dedicated Thomson servers to provide HA's

10 application, thus keeping a close commercial link with end users.

### 1.3.2. Typical interactions: some use cases

To illustrate how a user could interact with HA, below is a possible sequence of interactions with it (it assumes that some modules are loaded: "EPG", "Device Control", "Weather Forecast", "Encyclopedia").

User: "Any french comedy playing tonight?"  
[HA's EPG module scans the TV program and responds]  
HA: "Jour de fête, from Jacques Tati, will play at 11, tonight on Cinestar"

20           User: "Great! I don't have any copy yet. Record it for me, please"  
[HA's Device Control module sets up the TV receiver and video recorder to record the movie at 11:00pm]  
User: "Will I need my raincoat tomorrow?"  
[HA's Weather Forecast module retrieves the forecast from the Internet and informs the user]  
User: "Have I any appointment with my dental surgeon in the coming week?"  
[HA's Diary module consults the speaker's personal diary and answers accordingly]

25           User: "Retrieve the article on Schleswig-Holstein you showed me yesterday"

[HA's Encyclopedia module finds the article and displays it on the HA]

User: "Well, that's really a good one! Send it to Marylène!"

[HA's Encyclopedia module composes an e-mail with the article  
5 and sends it to Marylène]

### 1.3.3. Large Vocabulary, Spontaneous Speech Recognition

HA will recognize spontaneous speech (i.e. complete sentences, not only isolated words) with its associated large vocabulary.

10 It will also treat some hesitations (mumbling and silence), but not change of mind. From user tests we have conducted, we have observed that mumbling and silence cover more than 50% of all human speech hesitations.

15 Interactions with the HA will be possible via wireless microphones anywhere in the house, close-talking when standing in front of the HA, and also telephony (mobile or not) when away from home.

### 1.3.4. Speaker Dependency and Identification

When first bought, HA will use a user-independent recognition  
20 profile to cope with any possible user.

Users will be able to initiate an explicit training session, in order to increase the recognition rate in a significant manner. This higher recognition rate will be more and more useful with the number of different modules loaded in the HA.

25 In addition to increasing the recognition rate, this initial training will automatic, fine-grained, adjustment of the recognition rate for the identified user. It will also allow speaker identification, which is highly desirable to adapt the interactions to the particular user currently using the HA. This adaptation may even include particular modules that would be  
30 usable only by certain people in the home.

### 1.3.5.Text to Speech

The HA will interact visually through its display, but also vocally, through a Text-To-Speech interface. This interface will be available with multiple voice "styles", allowing a particular user to choose its preferred  
5 voice.

Of course, given the ability to identify the speaker (see above), each user at home will choose his/her own preferred voice for greater usage comfort.

10 1.3.6.Main Board

1.3.6.1.Platform

- A microprocessor,
- A RAM,
- A mass storage device for the Operating System, the

15 recognition engine and TTS feature plus the Home Assistant application, all grammars (~1-7 Mb each), associated semantic network, etc.

• Several Home Assistant microphone / display sets connected to the home network where is connected the Home Assistant main system.

20 1.3.6.2.Operating System

This may be a commercially available system used for personal computers.

1.3.6.3.Speech recognition engine

25 This may be a commercially available system used with personal computers.

1.3.7.Home Assistant Microphones

As we saw it previously, the Home Assistant system is not a mobile, the user must be able to use it in all the free space which he has, that is why the technique of the wireless microphone offers several important features :

5           • It allows the user to pilot the Home Assistant without obstacle and wherever it is with regard to the system,

              • It allows to localize the user in the space and there to offer automatically information or to supply wanted information there where is the user,

10           • Due to the technique of the " Close Talking " integrated in every mono-directional microphone, arranged in different places and offering each the same features, the problems associated to the " Distant Talking " can, at first, not be taken into account while authorizing this future feature in a real background noised environment.

15           The microphone does not inevitably have to be of very high quality, nevertheless its characteristics will mark the training phase and will determine there the rate of recognition of the system. A loss of 20 % of the recognition rate can be attributable in a bad use of the microphone or in an inadequacy between the microphone and the training voice model (bad  
20           pairing).

### 1.3.8.Feedback systems

Associated to the location of the user in its space due to microphones, the offer of contextual information or the supply of  
25           information asked by the user is made there where he places through one or several media following ones:

              • A display on one or several screens of identical or different characteristics (low resolution, high resolution, TV screen, etc.),

              • A vocal synthesis through one text to speech (TTS) engine,

30           • Any actions controlled by the Home Assistant (TV zapping, MP3 playing, etc.). See use cases of the Home Assistant.

#### 1.4. Software description

##### 1.4.1. Introduction : the problem to be solved

The Home Assistant must be so based on software architecture  
5 taking into account the notion of domains (TV, Internet, domotic, etc.) and  
to offer possibility for the user to move intuitively between several domains  
and to be able to return to one of them without losing the historic of route  
in this domain.

This can be obtained by developing a generic kernel including one  
10 or several voice recognitions engines and an manager of domains. The  
domain in that case can be represented as the set constituted with a  
specific semantic network, with a context, with a grammar and a typical  
vocabulary in the domain and, additionally, of the associated data.

The manager of domains in that case administers the load and  
15 dumping within the application of the complete sets - contexts, grammars,  
etc. - and the passage from a domain to the other one as at the level of the  
recognition engine (grammars, analyser, etc.) as elements not associated to  
the part voice recognition (Data base, display, TTS, etc.).

The other important part to develop is the tool of construction of  
20 these domains: it has to allow, by creating new capacities (new semantic  
network, etc.), to offer necessary features to cover varied domains of the  
Home Assistant system.

##### 1.4.2. Software Architecture

###### 25 1.4.2.1. Global Architecture

###### 1.4.2.2. The Context Manager

###### 1.4.2.2.1. Introduction

The context manager (CM), is responsible for handling a smooth  
dialog. The context manager analyses and stores at least the previous  
30 request and defines the context of the dialog. In order to further explain  
how the CM works, we first recall the three main steps of a dialog.

Then we show a few sample man-machine exchanges which illustrate the different kind of actions which must be performed by the CM.

We then classify those typical actions and give a first idea of the structures which have to be implemented in order to perform these actions.

5 Then, we show how the CM interfaces with the other Home Assistant modules with these structures.

Finally, we give a few further details about the internal structure of the CM.

10 1.4.2.2.2.The three main steps of a dialog

A dialog can be divided into dialog exchanges. Each dialog exchange is made of the following steps:

1. The user presents a demand to the system ;
2. The system evaluates the demand, extracting its meaning, verifying its coherence, and retrieving all the items which answer the demand ;
3. The system sorts the retrieved items in the appropriate order, chooses some of them and presents them to the user.

In this situation, the role of the context manager consists in the  
20 following :

1. It receives the user demand from the recognition module after this demand has been parsed and its meaning has been extracted ;
2. It analyses the demand to determine how it can be answered ;
3. It performs the appropriate actions to give the appropriate  
25 answer ;
4. It sends the answer to the Features Manager (FEM), for it to be displayed to the user.

The most important work of the context manager resides in step  
3. Hence, determining what are the appropriate actions to perform to give  
30 the appropriate answer to a demand is crucial to exhibit the internal structure of the CM.

#### 1.4.2.2.3.A few sample dialog exchanges

So, let's analyze the four typical dialog exchanges below:

5

1. User : "I'd like a movie".

The system answers.

User : "And what's after this" ?

2. User : "Is there a football match" ?

The system answers.

User : "And is there another one" ?

10

3. User : "What's on the first channel right now" ?

The system answers.

User : "And on the second one" ?

4. User : "I want a western please".

The system answers.

15

User : "I'd like to see cycling".

The system answers.

User : "Is there a tennis match" ?

The system answers.

User : "Could you show me the previous western" ?

20

#### 1.4.2.2.4.The main CM actions classification

In the first example, the user is considering the answer given by the system as a reference item and so, in his second question, he is asking for another item which is related to this reference item. In NLP terminology, 25 this reference item is called "the nominator" because it has the same "name" for both locutor. Therefore, in order to give an appropriate answer to the second user's question, the context manager must perform a "switch on the nominator".

The second example illustrates the fact that the user supposes the system is able to give not only one answer to his first question but all of them. In order to do so, the context manager must store them into a list we call "the stocker". In his second question, he is asking for another item of 5 this list. The appropriate action to be taken in this situation is called a "switch on the stocker".

In the third example, the user supposes that the system recalls his last demand. The place where the context manager stores this demand is called "the collectiviser" because it generally defines an ordered set of 10 items by giving its corresponding collectivising relation and the sort order in which the items of the set should be presented. In his second question, the user does not reformulate his whole demand (which should be "What's on the second channel right now"), but only the part of the collectivising relation which is different from the first one. Therefore, the appropriate 15 action which must be taken by the CM is called a "switch on the collectiviser".

In the last example, the user supposes that the system is able to recall previously displayed answers without having to reformulate the whole demand which generated them. In order to do so, the context manager 20 must record all the items which became the nominator at different times in what we call "the nominators". The action which must be performed by the CM to answer the last user question is called a "context callback".

Hence, the three main element we need to handle a smooth dialog are:

- 25 1. The most recent user demands ;
2. The last ordered set of items in which the system choose the most recent answer ;
3. The most recent answers which were given to the user.

These three elements form what we call "the context" of the 30 dialog.

In order to properly handle the context, the CM has the following structures:

- The collectiviser which contains a representation of the user demand ;
- 5           • The stocker which contains the possible answers to this demand ;
- The memory, (or nominators), which contains a list of items which were referred to by both the user and the system at different times.

#### 10           1.4.2.2.5.The CM interface

As the collectiviser contains a representation of the user demand, it is given to the CM by the recognition module.

The CM then converts the collectiviser into one or more complete and consistent requests. These requests are then sent one by one to a 15 specialised module called "the query manager".

For each request, the query manager fills the stocker with the item which satisfy it.

After all the requests are treated by the query manager, the stocker is sorted according to the order given into the collectiviser.

20           Finally, one or more items are chosen in the stocker and sent to the module specialised in presenting the answers to the user, the Features Manager (FEM). As soon as an item is sent to the FEM, it is stored in a "nominator" so that it can be recalled later by the CM.

Figure 3 is a graphical representation of the context manager 25 interfacing.

#### 1.4.2.2.6.The internal structure of the CM

The collectiviser is made of two kinds of data :

- The pertaining criteria which define the collectivising relation 30 itself ;

- The ascriptive criteria which define the sort order of the set which has to be produced in the stocker.
- Figure 4 is a graphical representation of the collectiviser.
- The stocker is represented by two lists :
  - The items list which contain the set of items which have been retrieved by the query manager ;
  - The ascriptive list which contain all the ascriptive criteria which have been extracted from the collectiviser and are used to sort the items list.

10                  Figure 5 is a graphical representation of the stocker.

The context manager memory stores each item independently as soon as it becomes the nominator. Hence, in this document, the term "nominator" further designates an individual item which is stored in the 15 context manager memory. As the memory cannot grow indefinitely, we choose to represent it as a list of items which has a fix size. When the list is full, each new item replaces the least recently used one. Hence, a nominator is represented with the following attributes:

- An individual item;
- A stamp which represents the last time and date at which the item was referred to in the dialog.

20                  Figure 6 is a graphical representation of a nominator.

#### 1.4.2.3.Interface to each Context : The Semantic Network

25                  SEMNET is an abbreviation for Semantic Network.

A SEMNET is a synthetic representation of the knowledge for one domain. With one SEMNET, we claim we can cover a large part of one domain. A domain is similar to an application (for example an EPG).

As shown in the global architecture, a domain is associated with a database, a SEMNET, a Context Manager and grammars. The database contains elements such as name of movie, name of actor, a day of the week, a town, an identifier of a document (the title), etc. These elements  
5 are associated with attributes which define the topic of the element, such as "actor", "sport", etc. A whole of selected attributes can define a domain. The domain contains among other things, the whole of elements of the database that have one of the selected attributes.

SEMNET insure the consistency between these entities.

10 As domains are very disparate, it's very important to define a generic way to formulate request and to implement event. SEMNET is the generic solution.

Here are the basics of SEMNET:

15           □ A criterion is a basic element. It can be an element of a database or an attribute.

          □ An event is an association of criterions. An event can define a whole of elements of the database that respect a list of attributes contained in a list of criteria.

20           □ Criterions are linked with relations.

#### 1.4.2.3.1.The SEMNET architecture.

Three basic relations are defined in SEMNET:

25           □ The Is\_A relation, see figure 9. This relation links an element of the database or at least one attribute of elements. For example, Woody Allen IS A actor, Woody Allen is an element of the database and actor is an attribute. Others links can exist, for example, in certain movies, Woody Allen is also a producer.

□ The Is\_AKindOf relation see figure 10. This relation links an attribute with another attribute. For example, "football" and "sport" are two attributes, the football is a sport.

5 □ The role (ROLE) see figure 11. This relation links two criteria that are equivalent, or in other word, synonymous. For example, a user can say as well "serial" and "movie" to designate the same concept

The list of these relations are not exhaustive.

10 These relations link different criterions.

As there is different way to join a criterion, SEMNET is a graph defining the elements of the domain and the links between these elements. When the user asks anything to the HA, the semantic network searches the answer in the graph to the request in the domain. If a criterion 15 corresponding to the request is reached, the HA or the semantic network work according to the status of the reached criterion.

A status is associated to a criterion. This status influences the behavior of the criterion during the execution of the request and the 20 searching of the answer.

Five basic status are defined in SEMNET:

25

- The Displayable Status.
- The Implicit Status.
- The Input Point Status.
- The Main Status.
- The Non Displayable Status.

Figure 7 is a representation of a SEMNET for EPG and Figure 8 is  
30 a view of a SEMNET for Cmd&Ctrl.

#### 1.4.2.3.1.1.The Is\_A relation.

The Is\_A relation allows the implementation of a Non Displayable criterion.

5

#### 1.4.2.3.1.2.The Is\_AkindOf relation.

The Is\_AkindOf relation increases the granularity.

#### 1.4.2.3.1.3.The Exclusive entity.

It is a variant of the Is\_A relation, illustrated by the figure 12. An  
10 exclusive entity allows only one selected criterion.

#### 1.4.2.3.1.4.The role.

An implicit criterion is attached to another criterion by a role. A role is an aggregation.

15

#### 1.4.2.3.1.5.The Status.

- The displayable status.

This is a non-abstract criterion (for example film is non-abstract).

Displayable criterions are directly searchable.

20

Displayable criterions are linked with Is\_AkindOf relations.

- The implicit status.

A role is ended by an implicit criterion. Implicit criterions aren't searchable alone.

25

Implicit criterions are linked with Is\_A relations.

- The Input Point Status.

These kinds of criterions are directly linked to the main criterion with an Is\_AkindOf relation.

30

In fact, this status defines a sub folder in a domain.

- The main status.

There is only one criterion with the main status. This one holds all the SEMNET.

The criterion with the main status holds all the master roles.

5

- The non-displayable status.

This is an abstract criterion and also an implicit criterion. Non-displayable criterions aren't searchable.

This criterion doesn't contain pertinent information.

10

#### 1.4.2.3.2.How does it work?

SEMNET implements methods for building the events and especially good requests.

A grammar is decorated by several points of generation. These 15 points reach criterions to compose requests or events.

An event is build with well-known information. On the other hand, the request is the result of grammar analysis. It's very difficult to ensure a good and non-ambiguous request. This is why SEMNET implements methods to solve these problems.

20

The major features are:

- SEMNET allows incompletely request.

For example, suppose the request " I want something with Julia Roberts".

Julia Roberts is only an actor of cinema.

25

The Database's engine searches in the cinema subfolder.

- SEMNET detects ambiguous requests.

For example, suppose the request " I want something with Woody Allen".

30

W. Allen is an actor of cinema but also an actor of theatre.

The returns of the request can be all the events with Allen as theatre's actor and as cinema's actor.

In fact, SEMNET allows interactive dialogs :"Allen as a theatre actor's or Allen as a cinema' s actors".

5

- SEMNET detects bad request.

Some wrongs requests could be generated, SEMNET filters them. The HA does not send any information in answer to the bad request. A variant consists in sending an interrogative question in answer to the wrong request.

10

#### 1.4.3.Creating and Developing a new context : user test loops

In order to develop a new context, our methodology is based on iteration loops: user tests - improvement of the Language Model - second user tests - etc.

We need at first a small LM, built according to what we think represents a minimal set of queries that should be admitted in this context. A first internal test allows us to improve this first LM and to make it more robust. It is then possible to build a mock up of the future application, which will be tested with external users. These tests will allow us the constitution of a linguistic corpus, with which we will define the corresponding semantic templates. A series of iterative loops (tests and improvement of templates and LM) will then be made until satisfaction of the users. This step is dependent on a given language.

25 See figure 13.

In order to create a new context, we have the following elements:

- The corpus collected during the user tests: WAV files,

- The results of recognition: XLS files, containing the questions made by users, the corresponding recognized sentences, the requests sent to the database and the responses given by the database,
- Our tool.
- 5           • With these elements, we proceed like shown in figure 14.
- For each sentence, there are two possibilities:
  - The sentence is recognized: the sentence is kept and we shall verify after evat each loop that it remains in the LM. This sentence will also allow us, after grouping between similar sentences, to build the
  - 10          corresponding semantic template.
    - The sentence is not recognized: two possibilities again:
      - \* the sentence is not included in the LM : it must be added in the LM or not, if it is out of context,
      - \* the sentence is included in the LM : it is an engine recognition's problem.

15

#### 1.4.4. 'No regression' requirement

After each loop, we re-inject the whole corpus of the previous steps (WAW format) in the system: all the sentences that were OK must remain OK. The evolution must be the one illustrated by figure 15.

20

#### 1.4.5. Language independence

The second step in our methodology consists in the internationalization. When the semantic templates are established, it is possible to build the LM for another languages than the language used in  
25          the initial step. Here again, we make some iterative loops (user tests and improvement of LM) until good results are obtained. See figure 16.

#### 1.4.6. HA Tool Software Architecture

The actual problem is that all the work described previously is "handcraft" made. The tool should help us to know, for each rejected sentence whether it is a LM problem or a recognition problem. It should also help us to verify if a recognized sentence is still recognized after each  
5 loop in the LM. See figure 17

## Claims

1. Stand-alone device (HA) containing a voice recognition system,  
5 a mass storage device ; characterized in that it comprises moreover a natural language processing (NLP), at least one semantic network (SEMNET) which defines a domain, a database containing information with attributes, the natural language processing (NLP) receiving recognized command of an user and sending a request to the semantic network (SEMNET), the  
10 semantic network being constituted as a graph defining the elements of the domain and the links between these elements, the semantic network searching the answer to the request in the domain.

2. Stand-alone device according to the claim 1, characterized in that the semantic network (SEMNET) comprises criteria which are the basic  
15 element of the graph, and relations which link the criteria; a criteria is an element of the database or at least one attribute of elements.

3. Stand-alone device according to the claim 2, characterized in that the semantic network (SEMNET) defines a relation between an element and at least one attribute (Is\_A).

20 4. Stand-alone device according to the claim 2, characterized in that the semantic network (SEMNET) defines a relation between an attribute with another attribute (Is\_AKindOf).

25 5. Stand alone device according to the claim 2, characterized in that the semantic network (SEMNET) define a relation between two criteria indicating that they are equivalent (ROLE),

6. Stand-alone device according to the claim 2, characterized in that a status is associated to a criterion, the status defining the behavior of the device or the semantic network when the request of the user reaches this criterion.

7. Stand-alone device according to the claim 1, characterized in that it comprises a context manager (CM), the context manager stores at least the previous request and informs the semantic network about the context of these previous requests.

5           8. Stand-alone device according to one of the previous claims, characterized in that the semantic network detects the bad request, and filtering them.

10          9. Stand-alone device according to one of the previous claims, characterized in that it comprises a means for identifying the user speaking to the device.

See Available Copy

1 / 8

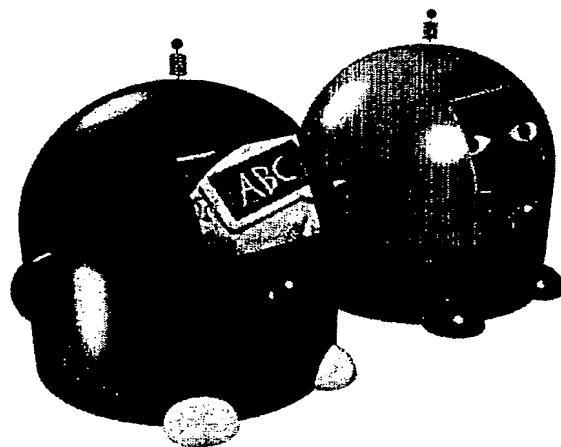


Fig. 1

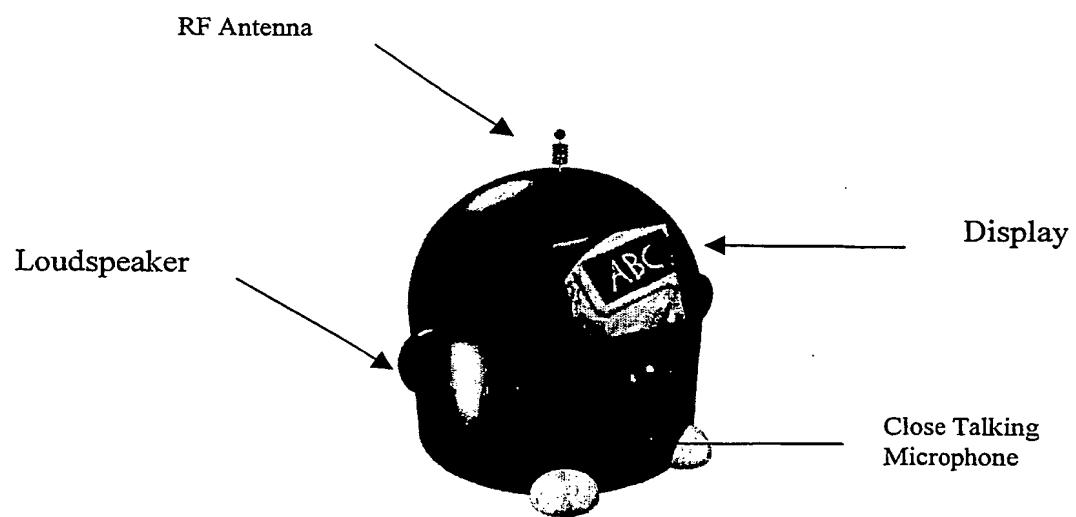


Fig. 2

Best Available Copy

2 / 8

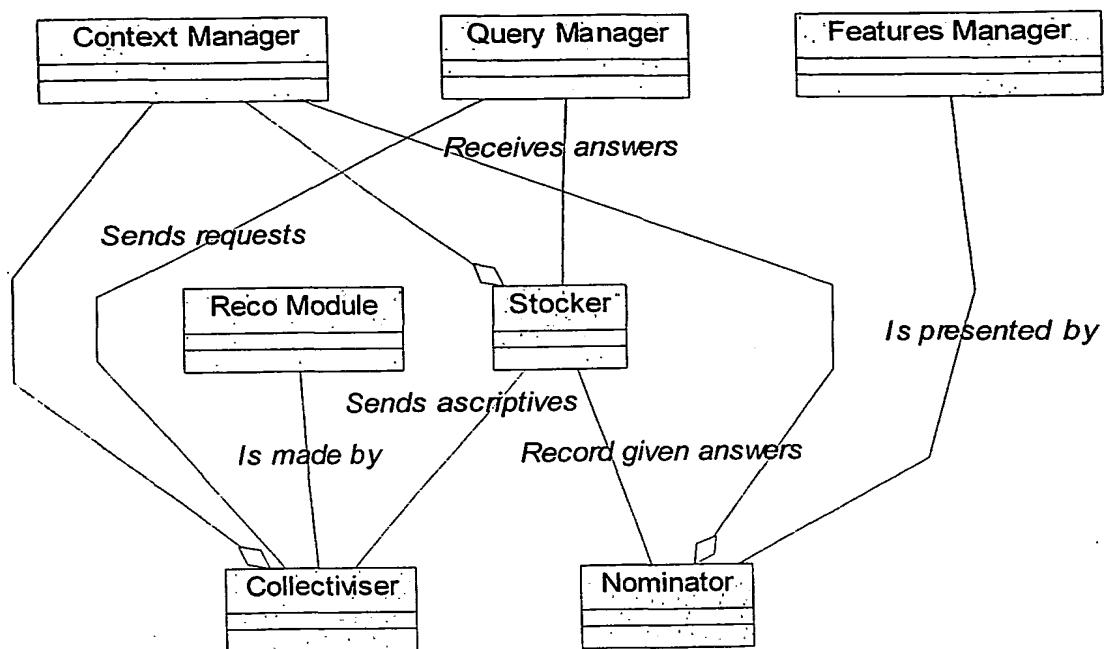


Fig. 3

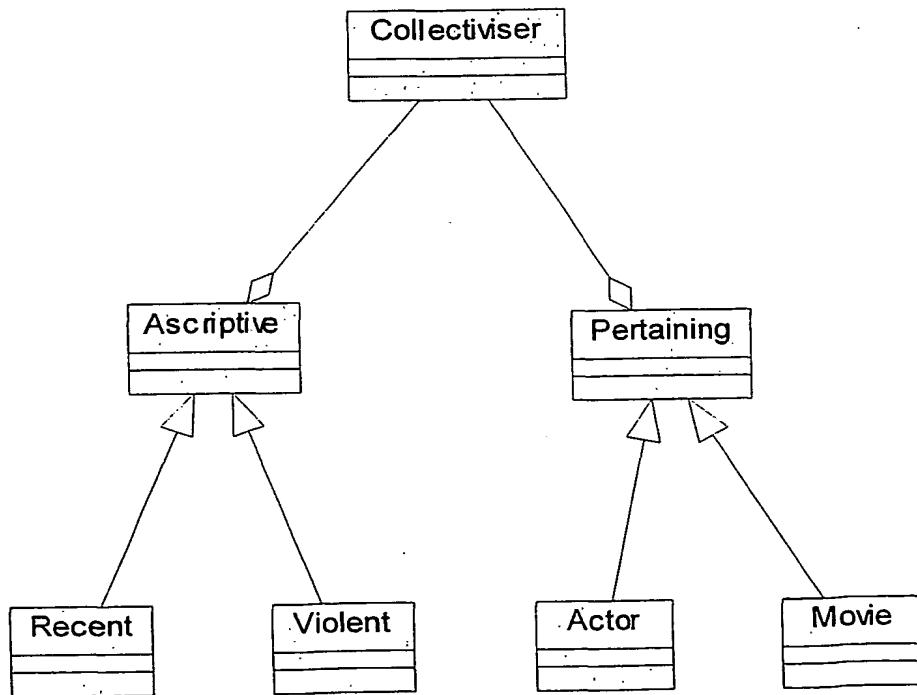


Fig. 4

3 / 8

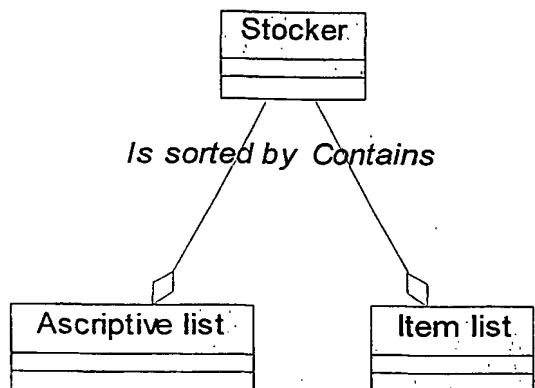


Fig. 5

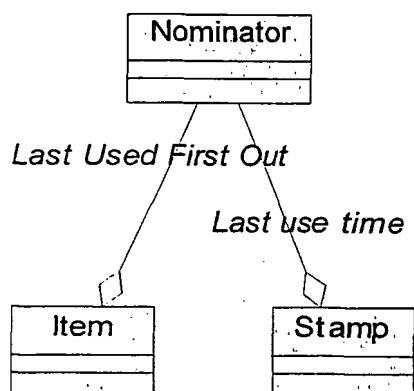


Fig. 6

4 / 8

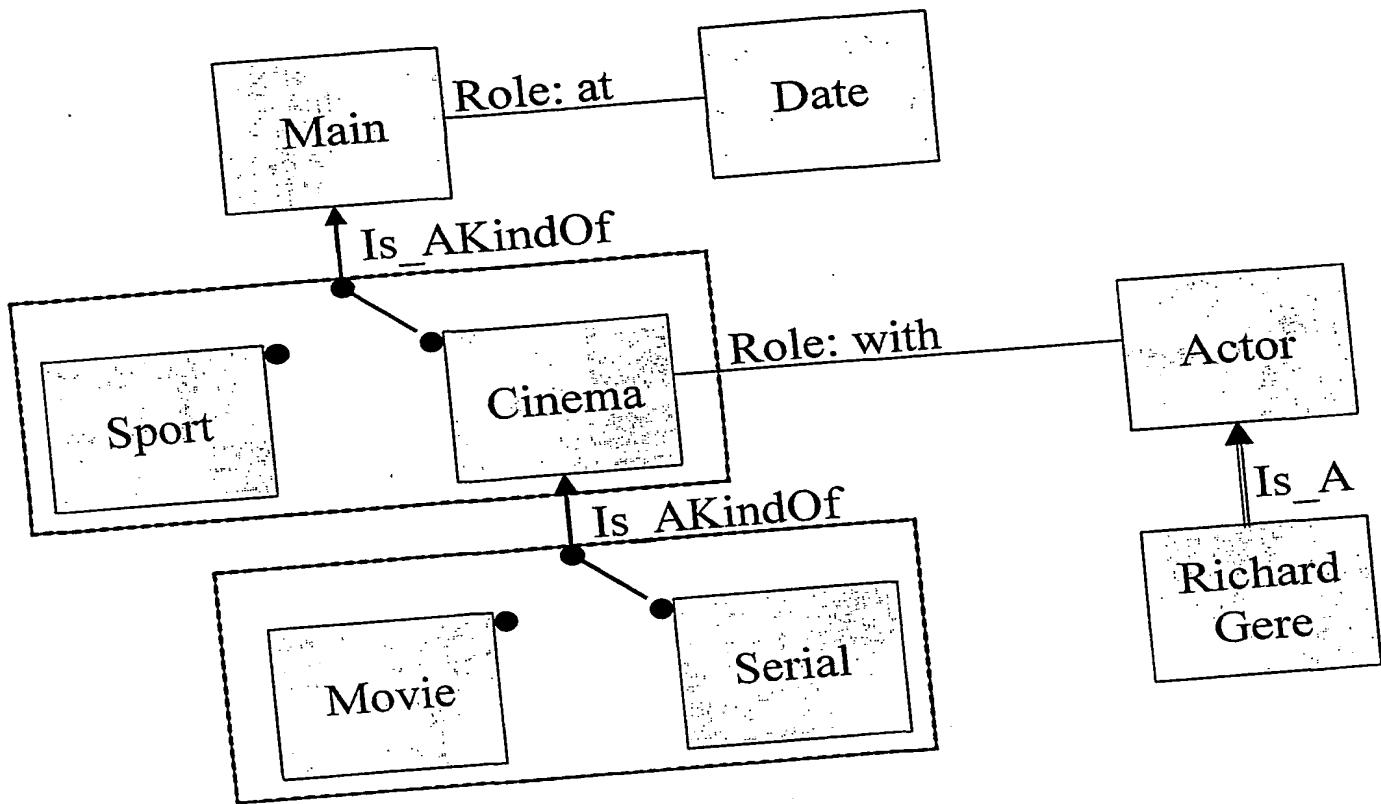
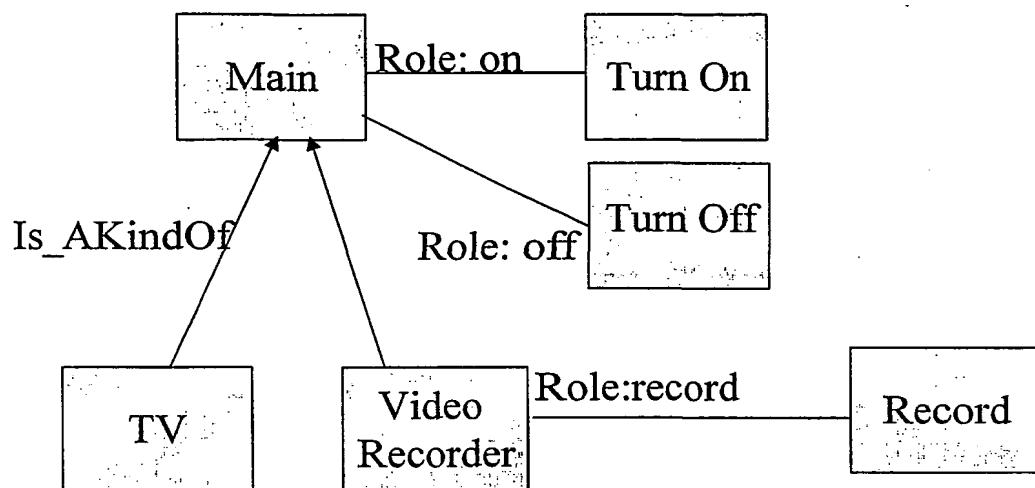
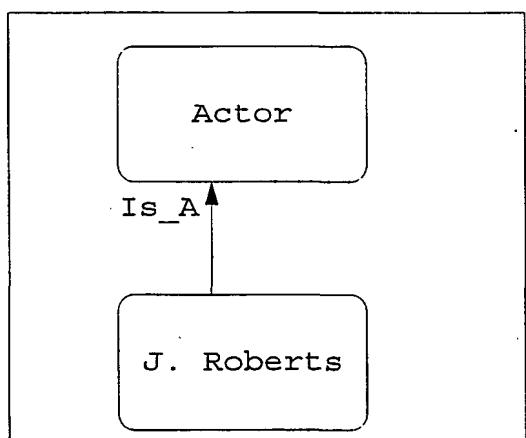


Fig. 7

**5 / 8****Fig. 8****Fig. 9**

6 / 8

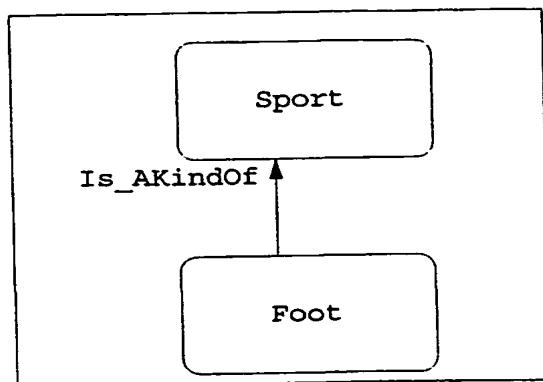


Fig. 10

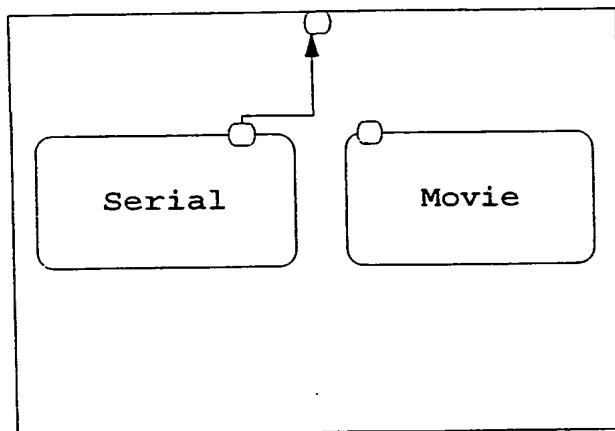


Fig. 11

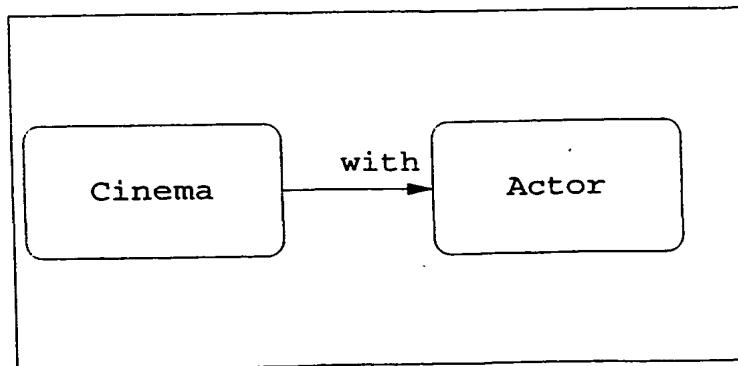


Fig. 12

7 / 8

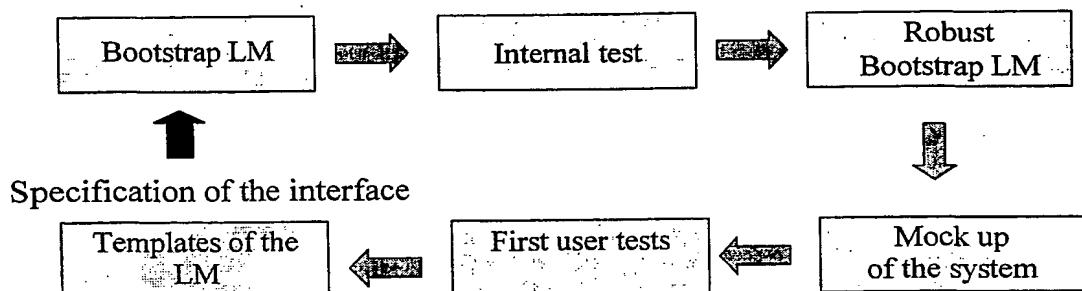


Fig. 13

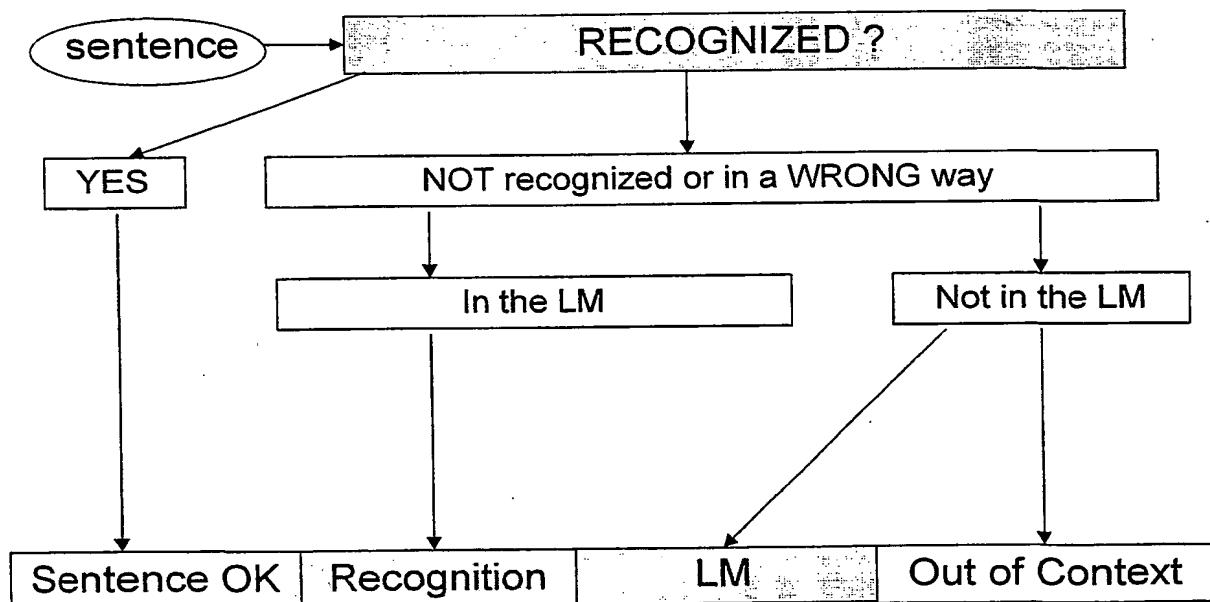


Fig. 14

STEP	Sentence OK	Recognition	LM	Out of Context
STEP				
STEP N+1:	Sentence OK	Recognition	LM	Out of Context

Fig. 15

8/8

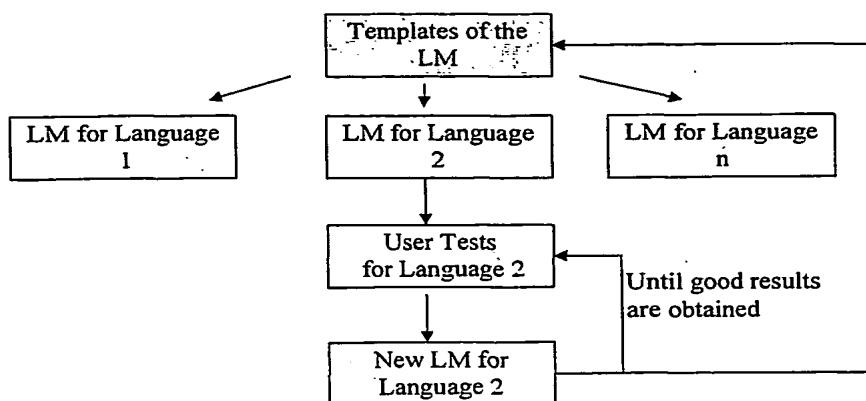


Fig. 16

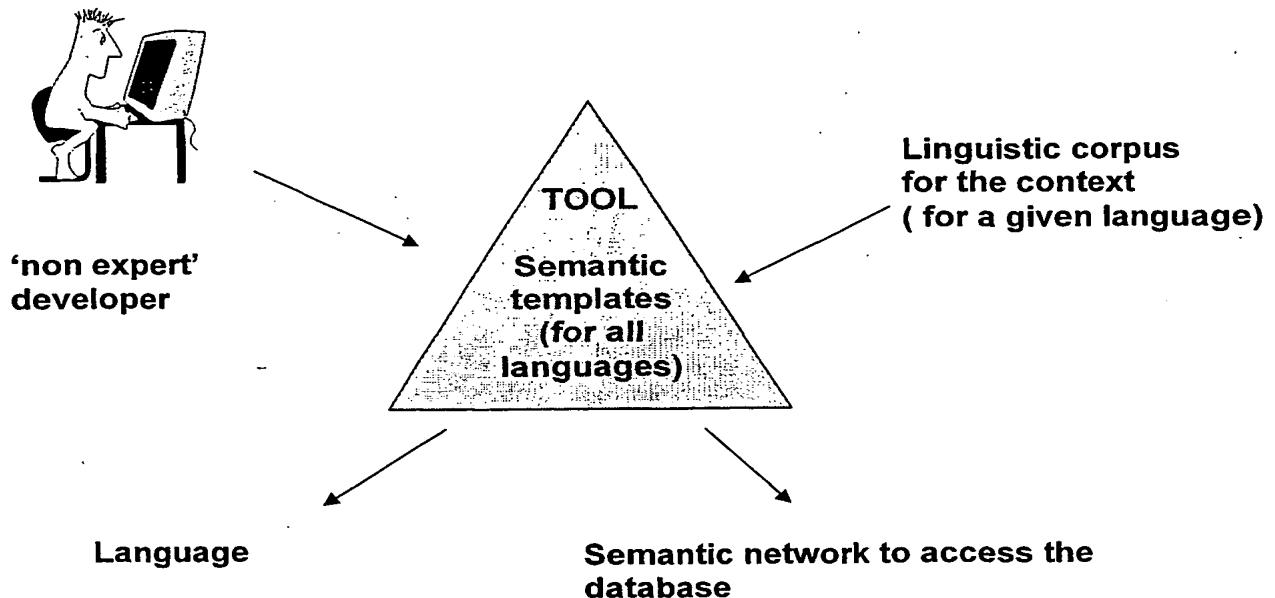


Fig. 17

## INTERNATIONAL SEARCH REPORT

national Application No

PCT/EP 01/05945

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G10L15/26

According to International Patent Classification (IPC) or to both national classification

## B. FIELDS SEARCHED

Minimum documentation searched (classification system used)  
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, PAJ, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MAST M ET AL: "A SPEECH UNDERSTANDING AND DIALOG SYSTEM WITH A HOMOGENEOUS LINGUISTIC KNOWLEDGE BASE"          IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE INC. NEW YORK, US,          vol. 16, no. 2,          1 February 1994 (1994-02-01), pages          179-193, XP000439829          ISSN: 0162-8828          page 182, right-hand column -page 183,          right-hand column          page 185, right-hand column -page 187,          left-hand column          ----          -/-</p>	1-3, 6-9

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

12 September 2001

21/09/2001

## Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

## Authorized officer

Ramos Sánchez, U

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 05945

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 18260 A (UNIV OXFORD BROOKES ; ZAJICEK MARY (GB); BROWNSEY KENNETH (GB)) 13 June 1996 (1996-06-13) page 2, line 1 - line 11 page 2, line 26 -page 3, line 10 page 5, line 8 - line 24 page 8, line 13 - line 27 page 11, line 25 -page 12, line 10 ---	1-9
A	EP 0 862 159 A (AGFA GEVAERT NV) 2 September 1998 (1998-09-02) page 5, line 38 - line 42 ---	9
A	GORIN A L ET AL: "How may I help you?" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 23, no. 1-2, 1 October 1997 (1997-10-01), pages 113-127, XP004117213 ISSN: 0167-6393 the whole document ----	1-9
A	KELLNER A ET AL: "PADIS - An automatic telephone switchboard and directory information system" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 23, no. 1-2, 1 October 1997 (1997-10-01), pages 95-111, XP004117212 ISSN: 0167-6393 the whole document -----	1-9

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/01/05945

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 9618260	A	13-06-1996	AU	4182596 A		26-06-1996
EP 0862159	A	02-09-1998	JP	10272106 A		13-10-1998
			US	6047257 A		04-04-2000